

500.43556X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): H. KUMAI

Serial No.: 10/790,129

Filed: March 2, 2004

Title: INFORMATION MINING SYSTEM

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

March 15, 2004

Sir:

Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby
claim(s) the right of priority based on:

Japanese Patent Application No. 2003-159274
Filed: June 4, 2003

A certified copy of said Japanese Patent Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP

Paul J. Skwierawski

Registration No.: 32,173

PJS/rr
Attachment



日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 6 月 4 日
Date of Application:

出 願 番 号 特 願 2 0 0 3 - 1 5 9 2 7 4
Application Number:

[ST. 10/C] : [J P 2 0 0 3 - 1 5 9 2 7 4]

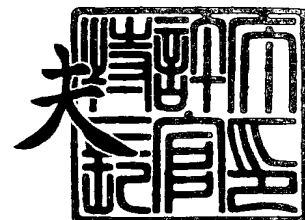
出 願 人 株式会社日立製作所
Applicant(s):



2 0 0 4 年 3 月 2 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 4 - 3 0 1 5 5 7 8

【書類名】 特許願

【整理番号】 GM0303105

【提出日】 平成15年 6月 4日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/30

【発明者】

 【住所又は居所】 東京都国分寺市東恋ヶ窪一丁目 2 8 0 番地 株式会社日立製作所 中央研究所内

 【氏名】 隈井 裕之

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社日立製作所

【代理人】

 【識別番号】 100075513

 【弁理士】

 【氏名又は名称】 後藤 政喜

【選任した代理人】

 【識別番号】 100084537

 【弁理士】

 【氏名又は名称】 松田 嘉夫

【選任した代理人】

 【識別番号】 100114236

 【弁理士】

 【氏名又は名称】 藤井 正弘

【手数料の表示】

 【予納台帳番号】 019839

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 0110326

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 情報マイニングシステム

【特許請求の範囲】

【請求項 1】

データベースに格納されたデータを、分析ツールを用いて分析する情報マイニングシステムにおいて、

前記分析において、過去に行われた分析ツールの操作履歴及び分析中の分析ツールの操作履歴を記録する操作履歴データ保持部と、

前記過去に行われた分析ツールの操作履歴と前記分析中の分析ツールの操作履歴との類否を判定し、該過去に行われた分析ツールの操作履歴と該分析中の分析ツールの操作履歴の少なくとも一部とが類似すると判定された場合に、前記分析中の分析ツールに対してその旨を通知する履歴比較部と、
を備えることを特徴とする情報マイニングシステム。

【請求項 2】

前記分析ツールの操作が入力され、前記分析ツールが備わっている端末が接続されており、

前記操作履歴データ保持部は、前記端末に入力された前記分析ツールの操作の履歴を記録することを特徴とする請求項 1 に記載の情報マイニングシステム。

【請求項 3】

前記端末は、前記データベースに格納されたデータの少なくとも一部を保持する分析用データ保持部を備え、

前記分析ツールは、前記分析用データ保持部に保持されたデータを分析することを特徴とする請求項 2 に記載の情報マイニングシステム。

【請求項 4】

操作履歴データが、前記操作履歴データ保持部に記録された操作履歴のいずれに記録されているかを対応付けて保持する帰結データ保持部を備えることを特徴とする請求項 1 から 3 のいずれか一つに記載の情報マイニングシステム。

【請求項 5】

前記帰結データ保持部は、前記分析によって得られた帰結を記録し、前記操作

履歴を該帰結との対応付けを保持し、

前記履歴比較部は、前記操作履歴データ保持部の記録内容と前記帰結データ保持部の記録内容の少なくとも一部とが類似すると判定された場合に、前記分析中の分析ツールに対してその旨を通知することを特徴とする請求項 4 に記載の情報マイニングシステム。

【請求項 6】

前記履歴比較部は、前記過去に行われた分析ツールの操作履歴と前記分析中の分析ツールの操作履歴の少なくとも一部とが類似すると判定された場合に、前記操作履歴に対応させて記録した帰結を分類することを特徴とする請求項 5 に記載の情報マイニングシステム。

【請求項 7】

前記履歴比較部は、

前記帰結及び前記操作履歴の双方が類似する第 1 の場合と、

前記帰結は類似しないが、前記操作履歴が類似する第 2 の場合と、

前記帰結は類似し、前記操作履歴が類似しない第 3 の場合と、

前記帰結及び前記操作履歴の双方が類似しない第 4 の場合とに、前記帰結データ保持部の記録内容を分類することを特徴とする請求項 6 に記載の情報マイニングシステム。

【請求項 8】

データベースに格納されたデータを、分析ツールを用いて分析する情報マイニングシステムに接続され、前記分析ツールの操作が入力される情報マイニング端末において、

前記情報マイニングシステムによって、分析中の分析ツールの操作履歴の記録内容と、前記分析において行われた分析ツールの操作履歴の記録内容の少なくとも一部とが類似すると判定された場合に、その旨の表示をすることを特徴とする情報マイニング端末。

【請求項 9】

前記分析ツールが備わっており、

前記情報マイニング端末に備わる計算リソースを用いて、前記データベースに

格納されたデータを分析することを特徴とする請求項 8 に記載の情報マイニング端末。

【請求項 1 0】

前記データベースに格納されたデータの少なくとも一部を保持する分析用データ保持部を備え、

前記分析ツールは、前記分析用データ保持部に保持されたデータを分析することを特徴とする請求項 9 に記載の情報マイニング端末。

【請求項 1 1】

データベースに格納されたデータを分析ツールを用いて分析する情報マイニングシステムに、該データを分析させるためのプログラムであって、

前記分析において行われた分析ツールの操作履歴と、該分析によって得られた帰結とを対応付けて記録する手順と、

分析中の分析ツールの操作履歴を記録する手順と、

前記帰結と対応付けられた操作履歴と、前記分析中の操作履歴との類比を判定する手順と、

前記帰結と対応付けられた操作履歴と、前記分析中の操作履歴の少なくとも一部とが類似すると判定された場合に、前記分析中の分析ツールに対してその旨を通知する手順と、を実行させるプログラム。

【請求項 1 2】

データベースに格納されたデータを分析ツールを用いて分析する情報マイニングシステムに接続され、前記分析ツールの操作が入力される情報マイニング端末に、該データの分析に関わる操作を実行させるためのプログラムであって、

前記情報マイニングシステムによって、分析中の分析ツールの操作履歴の記録内容と、前記分析において行われた分析ツールの操作履歴の記録内容の少なくとも一部とが類似すると判定された場合に、その旨の通知を受信する手順と、前記受信結果に基づいてその旨の表示をする手順と、を実行させるプログラム。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は情報マイニングシステムに関し、特に、分析者が効率よくマイニングを行える情報マイニングシステムに関する。

【0002】

【従来の技術】

近年、企業が保有する情報の電子化が進み、データウェアハウスに代表される大規模データベース構築手法を活用することで、より多くの数値、テキスト、図面等、電子化された情報が蓄積される。この蓄積された情報を含む大規模データベースを企業の業績に反映させるため、CRM (Customer Relationship Management)、SFA (Sales Force Automation) などの手法を取り入れる企業が増えている。

【0003】

これら大量の情報から有益な知見や新たなルールを抽出する手法として、情報マイニング技術が注目されている。

【0004】

この情報マイニング技術には、例えば、定量的な知見を得るデータマイニングがある。具体的には、相関ルール生成手段はデータベースを基に相関ルールファイルを生成し出力し、評価尺度指定手段はユーザが指定した評価尺度に基づき評価尺度ファイルを出力し、相関ルール評価手段は評価尺度ファイルに基づいて評価値を算出し、相関ルールファイル中の相関ルールの価値に関する情報を更新し、実行結果表示手段は相関ルールファイルに基づいて再評価された相関ルールの価値情報を表示すると共に、表示順序の並び替えや、表示する相関ルールの限定処理を行った上で相関ルールを表示することによって、ユーザの使用目的毎に異なる相関ルールの評価尺度を設定できるようになり、データマイニングの結果抽出された大量の相関ルールの中から、ユーザが自分の目的に役立つ有効な相関ルールを見つけ出す作業を容易に行うことができる（例えば、特許文献1参照。）

。

【0005】

また、定性的な知見を得るテキストマイニングも知られている。具体的には、言語特徴分析装置によって言語解析装置の精度向上のために文書データから分野

依存辞書を作成し、言語解析装置によって共起関係と係り受けを考慮した構文木を作成し、パターン抽出装置がこの構文木を用いることによって適切に頻出パターン（即ち、知識）を抽出・出力することによって、より細かく分類して知識抽出することができる（例えば、特許文献 2 参照。）。

【 0 0 0 6 】

【特許文献 1】

特開平 1 1 - 1 5 8 4 2 号公報

【特許文献 2】

特開 2 0 0 1 - 8 4 2 5 0 号公報

【 0 0 0 7 】

【発明が解決しようとする課題】

前述した従来の情報マイニング技術では、特に人間が記述したテキストを対象とする、マイニング技術を用いたツールによるマイニング作業は、知識発見の手順や、結果の評価などに、人間の高度な知識やノウハウが要求され、将来にわたって自動化が大変困難な業務である。このように高度な知識をもつ専門家の労働集約的な業務であり、専門家集団の育成、維持に多大な費用を要するため、企業での導入は進んでいないのが現状である。しかし、グローバル競争が激しくなる中、企業は内部に蓄積された知識を有効に活用できなければ、勝ち残っていけない状況にある。

【 0 0 0 8 】

この点に着目し、情報マイニング業務を企業から受託する情報マイニングサービスを提供する業者も現れ始めているが、マイニング作業は、属人的な作業であるため、複数の分析者の協調作業や、その効率化技術については考慮されていなかった。

【 0 0 0 9 】

本発明は上記の問題点に鑑みてなされたものであり、マイニングサービスを行う複数の分析者の協調的なマイニング作業において、重複した探索を防ぎ、効率よく多様な知識発見を行える情報マイニングシステムを提供することを目的とする。

【 0 0 1 0 】**【課題を解決するための手段】**

本発明の情報マイニングシステムは、複数の分析者が同時並行的にマイニングを進める際に、相互の分析ツールの操作履歴データを逐次比較し、現在の操作履歴が、既に他の分析者によって帰結を得た履歴と類似していると判定した際に、重複した分析を行っている旨の警告を発するよう構成した。

【 0 0 1 1 】**【発明の実施の形態】**

以下に本発明の実施の形態を図面を用いて説明する。

【 0 0 1 2 】

図 1 は、本発明の第 1 の実施の形態の情報マイニングシステムを中心とした構成を示すブロック図である。

【 0 0 1 3 】

1 は情報マイニングシステムであり、例えば、サーバ等のコンピュータによって構成される。この情報マイニングシステム 1 は、インターネット等のネットワーク 5 - a、5 - b、5 - c を介して、情報マイニングサービスを依頼する依頼者のコンピュータシステム 2、情報マイニングシステムの管理を行う受託者のコンピュータ 3、データベースのマイニングを行うクライアントとして機能する分析者コンピュータ 4 に接続されている。

【 0 0 1 4 】

依頼者のコンピュータシステム 2 には、情報マイニング対象のデータを電子データベースとして格納した分析対象データベース 6 が備えられている。この分析対象データベース 6 は、データウェアハウス技法等により構築されており、例えば、コールセンターにおける顧客との対応履歴（一般に自由文で格納されている）、顧客の属性データ（一般にコード化された数値で格納されている）、アンケートデータと回答者の属性、営業報告と社員の属性等のデータである。

【 0 0 1 5 】

情報マイニングシステム 1 には、分析用データ生成部 1 1、分析用データ保持部 1 2、一般情報データ保持部 1 3、分析ツール 1 4、操作履歴データ保持部 1

5、帰結データ保持部16、履歴比較部17等が備えられている。

【0016】

依頼者の分析対象データベース6は、ネットワーク5-a等を介して分析用データ生成部11に送られ、分析用データ生成部11で情報マイニングシステム1によって処理可能なデータベース形式に変換され、分析用データ保持部12に保存される。分析者は、分析者コンピュータ4を使用して、分析ツール14を操作する。

【0017】

分析ツール14は、分析用データ保持部12と一般情報データ保持部13に保存されたデータを参照し、分析者の操作の履歴を操作履歴データ保持部15に保存する。分析中の操作履歴は、操作履歴データ保持部15の分析中操作履歴テーブルに一時的に格納されている。そして、帰結を登録すると、操作履歴は帰結と対応付けられ、操作履歴データ保持部15の操作履歴テーブルに記録される。また、分析ツール14は、分析者によって登録された結果である帰結を帰結データ保持部16に保存する。この操作履歴及び帰結は、履歴比較部17によって内容が比較され、分析者コンピュータ4に比較内容が通知される。この操作履歴データ保持部15及び帰結データ保持部16は、メモリ、磁気ディスクなどの物理的な電子データ記憶媒体に、論理的な構造を持つデータベースとして構成される。

【0018】

なお、一般情報データ保持部13に保持されている一般情報データは、分析対象データベース6の情報にはない社会一般的な情報（例えば、ニュース、新聞、Web等の時系列に対応した時事の情報）である。この一般情報データは、分析ツール14によって利用可能な形式にされている。分析者は分析用データと一般情報データとを比較することで、分析結果と社会情勢との関連を見い出すこともできる。

【0019】

次に、第1の実施の形態の情報マイニングシステムの動作を説明する。

【0020】

従来、マイニングサービスの受託者は、分析者が登録した帰結の一覧から、帰

結のリスト及び分析の結果得られた知識のリストを参照していた。また、従来のマイニングサービスによる分析は、単独の分析者によって行われることが多く、帰結のリストに重複する帰結が現れる可能性はほとんどなかった。これは、情報マイニングシステムを用いた分析ニーズが限られていたためである。

【 0 0 2 1 】

しかし、近年、分析ニーズ、中でも迅速な分析ニーズの高まりに伴い、複数の分析者が同一の分析対象データを同時並行的に分析を行う必要性が生じている。特に、依頼者からの受託による情報マイニングサービスでは、時間的、費用的に、いかに効率よく、いかに多様で価値の高い知識を発見できるかが、依頼者からの高い評価、高い対価を獲得するための情報マイニングサービスビジネスのポイントとなる。

【 0 0 2 2 】

そこで、複数の分析者によるマイニングを行う方法が考えられる。しかし、複数の分析者が同時に分析する場合、他人が既に報告した帰結を導き出す分析を行ってしまうことがある。複数の人間により同じ帰結が得られたという事実は、その帰結や履歴の信頼度、すなわちその帰結や履歴の価値を高めることにはつながるかもしれないが、一定の期間で多様な情報を取得するという目的には適さない。

【 0 0 2 3 】

これを防ぐため、ある分析者から新たな帰結が報告された場合、該帰結を他の分析者に順次提示する方法が考えられる。しかし、この方法では、現在分析中の内容との関係に係わらず、他の分析者が得た帰結が提示されるため、他の分析者は帰結が報告されるたびに、その帰結の内容を確認するために、作業を中断させられる恐れがあり、分析者の集中力を削ぎ、結果として分析効率を低下させる可能性が高い。

【 0 0 2 4 】

そこで、本発明の情報マイニングシステム 1 では、他の分析者に対して以下に説明するように帰結を提示する。

【 0 0 2 5 】

図 2 は、第 1 の実施の形態の情報マイニングシステムの処理のシーケンス図である。このシーケンス図を次に示す 8 つのステップに分けて説明する。なお、図 2 では受託者は情報マイニングシステム 1 の管理者であり、受託者と情報マイニングシステムは一つに示してある。

【 0 0 2 6 】

まず、第 1 のステップでは、依頼者が情報マイニングシステムの受託者に対して、データマイニングを要求して、マイニングを委託する。

【 0 0 2 7 】

次に、第 2 のステップでは、依頼者が受託者に対して分析対象データベース 6 に保持されたデータの情報を通知して、情報マイニングシステム 1 から分析対象データベース 6 に接続可能にし（シーケンス 1）、分析対象データベース 6 のデータ内容をネットワーク 5 - a を経由して情報マイニングシステム 1 に転送する。また、データの転送が認められない場合には、依頼者のコントロールの下ネットワーク 5 - a を介した、情報マイニングシステム 1 から分析対象データベース 6 への接続を確立し、分析対象データへのアクセスを可能とする。なお、記録媒体に記録した分析対象データを依頼者から受託者に受け渡してもよい。

【 0 0 2 8 】

次に、第 3 のステップは、情報マイニングシステム 1 の分析用データ生成部 1 1 によって、分析対象データベース 6 の分析対象データを分析ツール 1 4 で使用可能なデータ形式に変換し、分析用データを生成して、分析用データ保持部 1 2 に保存する。

【 0 0 2 9 】

次に、第 4 のステップは、受託者が、複数の分析者（分析者コンピュータ 4）に対して、分析 I D（分析ツール 1 4 へのアクセスに関する情報、例えば、分析ツール 1 4 を使用するためのユーザ I D 等）を通知し、分析用データのマイニングを依頼する（シーケンス 2）。なお、分析者（分析者コンピュータ 4）は複数でなくてもよい。

【 0 0 3 0 】

次に、第 5 のステップでは、分析データのマイニング依頼を受けた分析者が、

分析者コンピュータ 4 を用いて分析ツール 14 に対して分析 ID をログインコマンドとして入力することで、情報マイニングシステム 1 にログインする（シーケンス 3）。分析 ID を入力することで分析ツール 14 が利用可能となり、まず、分析者コンピュータに分析ツール 14 の初期画面を表示する（シーケンス 4）。

【0031】

各分析者は、分析データをブラウズ（参照）し、何らかの仮説を立案して、分析ツール 14 を使用して仮説を検証する。この分析ツールは、後述するように、分析を様々な角度から分析するツールの集合であり、複数のツールから構成されている。

【0032】

分析者は、分析ツールを用いて分析を行う際に、分析ツール 14 のメモ欄にメモを電子的に書き留め、記録することができる。このメモは、分析ツールの処理の履歴と共に情報マイニングシステム 1 の操作履歴データ保持部 15 に逐次記録される。さらに、仮説の検証の結果やコメントをメモ欄に記入し登録することができる。例えば、仮説の検証に成功し何らかの有益な知識が得られた場合にはその知識を登録し、仮説の検証に失敗した場合にはその失敗を登録し、検証をあきらめた場合にはあきらめた事実を登録する（シーケンス 5、6、9）。

【0033】

分析者の分析作業中は、分析者が分析ツール 14 に対して行った操作の履歴が、分析者毎に、そして分析処理が行われる毎に、情報マイニングシステム 1 の操作履歴データ保持部 15 に逐次記録される。また、分析者が仮説の検証の結果を記入したメモを、帰結として登録することを指示した場合には、当該メモが「帰結」となる。情報マイニングシステム 1 は、帰結の内容を、帰結に至った一連の検証の操作履歴と対応付けて帰結データ保持部 16 に記録する。

【0034】

なお、帰結の登録が行われるまで、分析中の操作履歴は、操作履歴データ保持部 15 の分析中操作履歴テーブルに一時的に格納されている。そして、帰結を登録すると、操作履歴は帰結と対応付けられ、操作履歴データ保持部 15 の操作履歴テーブルに記録される。操作履歴テーブルに記録されると、対応する操作履歴

は分析中操作履歴テーブルから削除される。つまり操作履歴は移動される。

【0 0 3 5】

次に、第6のステップでは、情報マイニングシステム1は、現在分析中の分析者が分析操作を行う都度、すなわち、操作履歴が追加される都度に、履歴比較部17によって、分析者の現在の分析に関わる操作履歴と、操作履歴テーブルに記録されている過去の操作履歴とを比較する。履歴比較部17は、分析者の分析中の操作履歴が他の分析者（又は、分析者自身が過去に行った分析）によって登録された帰結に対応する操作履歴（又はその一部）と類似度が高いと判定した場合、分析ツール14（分析者自身の分析者コンピュータ4）に対して警告を通知すると共に、類似度が高い操作履歴に対応する帰結やメモの一覧等を表示する（シーケンス7）。この警告を受けた分析者は、警告に対する確認操作を行う（シーケンス8）。

【0 0 3 6】

分析者は、一覧を確認し自分が立案した仮説に対する帰結が、既に他の分析者（又は自分が過去に行った分析結果）によって得られているときは、重複した探索を防ぐために作業を中断することができる。

【0 0 3 7】

また、帰結が既に得られている場合においても作業を続行することもできる。これは、提示された時点までの操作履歴が類似していても、以降の分析で異なる帰結を得られたり、新たな知識の発見につながる事象を分析者が見出す可能性もあるからである。この場合、仮に同じ帰結に至った場合には、情報マイニングの多様性の点からは価値が低いが、帰結の確度を補強する情報となる。この履歴比較部17による類似度の判定は、分析対象の類似性、分析条件（検索キーワードや数値）の類似性、操作および操作の流れの類似性を数値化することによって求めることができる。

【0 0 3 8】

前述した第5及び第6のステップは繰り返し行われる。そして、第7のステップでは、予め定められた一定の期間（例えば、依頼者の指定した納期）が経過した場合、又は、予め定められた帰結の数（例えば、依頼者の指定したコストに基

づく成果) が得られた場合に、分析者は分析を終了し、分析ツール 14 からのログアウト処理を行う。この分析者のログアウト又は分析者からの終了の通知を契機として一連の分析処理が終了する (シーケンス 10)。

【0039】

そして、受託者は分析者の分析結果が報告した帰結の一覧を帰結データ保持部 16 から得る。この帰結の一覧は受託者が取りまとめる。この取りまとめは、例えば、表記や表現は異なるが同じ内容を示すと考えられる帰結を一つにまとめたり、得られた帰結の内容を勘案して評価を行う。

【0040】

次に、分析者が操作する分析ツール 14 の概要について説明する。第 1 の実施の形態の分析ツール 14 は、情報を様々な角度から分析するソフトウェアの集合であり、複数のツールからなる。なお、本発明の実施の形態では、情報のうちテキスト情報を分析するテキストマイニングツールを用いた例を示すが、OLAP ツールなどのデータマイニングツールも分析ツールの一つとして利用することが可能である。

【0041】

図 3、図 4、図 5 に、第 1 の実施の形態の、テキストから情報をマイニングする分析ツールの一例を示す。

【0042】

図 3 は、関連語を表示する関連語表示ツールの画面表示の例である。この関連語表示ツールは、分析対象データ中の文章中にどのような単語が頻出しているか、あるいはそれらの単語の一つ以上を指定して、同じ文書、あるいは同一の文、あるいは隣接して現れる単語にはどのようなものがあるか、その頻度はどの程度かなどの共起関係を提示できる。図 3 に示す例では、「ヒューズ」「配線」という単語を含む文章にはどのような関連語が表れているかを検索した結果であり、「切れる」、「交換」、「不良」、「切断」という結果が提示されている。

【0043】

図 4 は、指定した文字列を含む文章を検索する全文検索ツールの画面表示の例である。全文検索ツールは、例えば、関連語表示ツールで得られた関連語を含む

文書の一覧を参照することができ、関連語がどのような文脈で使用されているのかを直接確かめることができる。図4に示す例では、「ヒューズ」、「切れる」という単語を含む文章を検索した結果が提示されている。

【0044】

図5は、指定した文章と類似する文章を検索する類似文検索ツールの画面表示の例である。類似文検索ツールは、例えば、全文検索ツールで得られた文書の中から、一つの文章を指定し類似する文章の一覧を得ることができる。例えば、「PC」、「パソコン」、「パーソナルコンピュータ」といった表記の違いを同義語辞書を用いて吸収したり、指定した文章中の特徴的な語を抽出し、同じような特徴語を持つ文書を優先的に提示するなどの機能を持つ。全文検索では一覧に含まれないが、内容的に類似する文章を表示したいときに有効な分析ツールである。図5に示す例では、「アイロンのヒューズが切れたときにどのように交換すればよいか」という文章に対する類似する意味の類義文を検索した結果が提示されている。

【0045】

なお、これらの分析ツールには、メモ入力欄と帰結登録ボタンが備えられている。メモ入力欄は、コメントや帰結を入力する領域である。帰結登録ボタンは、一連の分析から何らかの帰結が得られた場合に、分析者がメモとして入力したコメントを帰結として登録する処理を指示するボタンである。

【0046】

次に、第1の実施の形態の情報マイニングシステム1の履歴データ保持部15の中の分析中操作履歴テーブル及び操作履歴テーブルについて説明する。

【0047】

図6に分析中操作履歴テーブルの一例を示す。分析中操作履歴テーブルには、記録内容の時系列的順序を保持する記録番号501、分析用データ保持部12中の操作の対象となったデータの内容を示す対象データベース識別情報502、使用した分析ツールの種別を示す分析操作種別503、当該操作における分析ツールで行った処理の内容を示す分析ツールに対して送ったパラメータ504の各フィールドを持つ。分析中操作履歴テーブルは、分析者毎（又は、ログイン単位）

に独立して用意され、現在分析中の操作履歴が一時的に格納される。

【0048】

図7に操作履歴テーブルの一例を示す。操作履歴テーブルには、記録内容の時系列的順序を保持する記録番号601、分析用データ保持部12中の操作の対象となったデータの内容を示す対象データベース識別情報602、使用した分析ツールの種別を示す分析操作種別603、当該操作における分析ツールで行った処理の内容を示す分析ツールに対して送ったパラメータ604、当該操作がどの分析者によって行われたかを示す分析者識別情報（分析者ID）605の各フィールドを持つ。

【0049】

なお、分析中操作履歴テーブルの記録番号501は、操作履歴テーブルに移動される際に、帰結単位での時系列を保持したまま、操作履歴テーブルの記録番号601で一意な値を持つように、情報マイニングシステム1によって番号が振り直される。また、分析中操作履歴テーブル及び操作履歴テーブルのパラメータ503、603には、分析者が分析の任意の中間段階で作成したメモを、コメントとして格納できるようになっている。帰結やコメントは、テキストに限るものではなく、音声、イメージや動画などそれを参照する人が理解できるものであればデータの形式を問わないが、本実施例では、テキストで記録することを前提として説明する。

【0050】

次に、情報マイニングシステム1の帰結データ保持部16に保存される帰結テーブルについて説明する。

【0051】

図8は帰結テーブルの一例を示す。帰結テーブルは、帰結を識別する一意の識別子である帰結番号701、当該帰結を登録した分析者を示す分析者識別情報（分析者ID）702、当該帰結に対応する操作履歴を前述した操作履歴テーブルの記録番号601の範囲で示す対応履歴番号703、帰結の内容として入力されたメモを示す帰結内容704の各フィールドから構成される。

【0052】

次に、履歴比較部 1 7 における類似度の判定処理について説明する。

【 0 0 5 3 】

図 9 は、履歴比較部 1 7 で行われる操作履歴の類似度を判定するフローチャートである。

【 0 0 5 4 】

分析者が分析ツールの操作を実行すると、まず、当該分析者が現在行っている分析作業の操作履歴を、分析中操作履歴テーブルから取得する（ステップ 1 0 0 1）。

【 0 0 5 5 】

次に、帰結テーブルと操作履歴テーブルを参照し、帰結テーブルの対応履歴番号によって対応付けられた一群の操作履歴を取得し、分析中操作履歴テーブルから取得した操作履歴との類似度を比較する（ステップ 1 0 0 2）。

【 0 0 5 6 】

このステップ 1 0 0 2 における類似度の算出には様々な手法がある。例えば、時系列的に分析の最初から予め設定された数の操作の対象とするデータベース、操作種別、パラメータ等が完全に一致した場合は、類似度は 1 0 0 % である。また、パラメータのみが違う場合には、その違いを類似度として評価することができる。なお、この予め設定された数の操作は、分析者によって行われた操作がある程度蓄積された後に類似度を判定するように、初期値を設定する。これは、操作の初期段階では多くの分析の処理が似かよっているため、ほとんどの操作が類似と判定されてしまうのを防ぐためである。

【 0 0 5 7 】

より具体的には、類似文検索ツールではパラメータの文章同士を、形態素解析の結果、形態素の同じ並びが存在し、その形態素数の全形態素に対する割合を類似度とすることができる。また、従来技術である特許文献 2 に示されているように文章から形態素を切り出し、係り受け関係（例えば、「ヒューズ」→「切れる」といった言葉の係り受け）を抽出し、その一致個数の、双方の文の合計文字数に対する割合を正規化したものも類似度として使用することができる。また、全文検索においては、パラメータの検索文字列を、「パソコン」と「パーソナルコ

ンピュータ」といった同義の文言の違いをDPマッチング手法で比較した結果を類似度とすることができる。さらに、データマイニングにおいては、パラメータのデータ同士の相関係数などを類似度として用いることもできる。操作履歴の比較毎に計算された類似度は、比較した操作数によって正規化され、トリガー値と比較される。

【0058】

そして、この類似度を予め定められた閾値（トリガー値）と比較判定する（ステップ1003）。比較判定の結果、類似度がトリガー値以上であると判定した場合、類似度が高いので、ステップ1004に進み、当該帰結を一時的なバッファである通知テーブルに記録する。トリガー値に満たないと判定した場合、類似していないので、当該帰結を通知テーブルに記録する（ステップ1004）ことなく、ステップ1005に進む。

【0059】

ステップ1005では、類似度を判定していない帰結テーブルの内容（帰結データ）が他にあるかを判定する。類似度の判定を行っていない帰結データがあればステップ1002に戻り、帰結テーブルに登録された帰結の数だけ類似度の判定を繰り返す。

【0060】

次に、前述した通知テーブルに帰結データが記録されているか否かを判定する（ステップ1006）。通知テーブルに帰結データが記録されている場合はステップ1007に進み、分析者に類似している旨の警告と共に、類似していると判定された帰結データの一覧を分析者コンピュータ4に通知する。一方、通知テーブルに帰結データが記録されていない場合は、既に登録された帰結データには分析者の操作内容に類似するものがないので、分析者に類似している旨の警告を通知（ステップ1007）することなく、そのまま処理を終了する。

【0061】

次に、図9のステップ1007における警告の通知について説明する。

【0062】

図10は、警告の通知時に分析者コンピュータ4に表示される画面の例である

。

【0 0 6 3】

警告は、分析者が分析ツール 1 4 の操作を実行した際に行われる履歴比較部 1 7 による類似度の判定結果に基づいて、分析ツール 1 4 によって分析者コンピュータ 4 に通知される。すなわち、履歴比較部 1 7 の結果に基づいて分析ツール 1 4 に対して通知される（分析者コンピュータに表示される）警告の表示が、警告通知部を構成する。

【0 0 6 4】

この警告を分析者に提示する警告ウィンドウ 8 0 1 には、警告の内容を示す警告メッセージ 8 0 2 と、類似する操作履歴の帰結の内容を示す帰結リスト 8 0 3 が表示される。この帰結リスト 8 0 3 には、帰結番号、帰結を登録した分析者の識別情報（分析者 I D）、帰結内容が含まれている。

【0 0 6 5】

なお、帰結番号をアンカーとし、分析者が帰結番号を指示すると対応する操作履歴を表示したり、分析者 I D をアンカーとし、分析者が分析者 I D を指示すると当該分析者の帰結一覧を表示する機能を、分析ツールに加えることができる。また、警告の画面とは別に、帰結リストを任意の時点で表示し参照する機能を分析ツールに加えることができる。このような追加機能は、W e b アプリケーションなど既知の技術で容易に実現可能である。

【0 0 6 6】

次に、第 8 のステップとして、分析者によって導き出された帰結によって受託者が行うサービスについて説明する。

【0 0 6 7】

図 1 1 は、分析された結果を受託者が依頼者に対して提示する処理を示すシーケンス図である。

【0 0 6 8】

前述したように、受託者は、分析者によって分析された結果（帰結及び帰結に至った履歴）を取りまとめ、各帰結に対する評価を行う。

【0 0 6 9】

そして、受託者は、依頼者の要求に応じ分析の経過で得られた操作履歴データを依頼者に提供するサービスを行う。帰結データは形式化された知識としての価値があるが、履歴データはその知識を得るノウハウなどの暗黙的な知識の一部を表出したものと捉えることができ、帰結に劣らない価値が存在していると考えられるからである。

【 0 0 7 0 】

受託者は、評価付けされた帰結のリスト（一覧）を依頼者に提示する。この際、依頼者のマイニング要求に対する帰結のみをリストに含め、マイニング要求の結果ではないが有用と思われる帰結を導き出した場合には、これを別途対価と引き換えに要求者に提示することもできる。

【 0 0 7 1 】

次に、受託者は取りまとめた帰結、帰結に対する評価、及び、帰結に対応する履歴を購入した場合の金額をそれぞれ提示する（シーケンス 1 1）。

【 0 0 7 2 】

依頼者は、帰結を参照して、帰結に対する履歴が必要だと判断した履歴の購入を決定すると、その履歴を履歴購入対象リストとしてまとめ、購入依頼を行う（シーケンス 1 2）。

【 0 0 7 3 】

受託者は、履歴の購入の依頼を受け付けると、履歴の購入対価の金額を計算し、依頼者に提示し請求する（シーケンス 1 3）。依頼者はこれを受けて、対価の支払いの承諾を得て、対価の支払い承諾通知を受託者に送る（シーケンス 1 4）。

【 0 0 7 4 】

受託者は支払受諾通知を受け、課金処理を行い、対象となる履歴データの開示処理を行う（シーケンス 1 5）。依頼者はこれを受けて購入した履歴を参照できる。

【 0 0 7 5 】

なお、分析ツール 1 4 の利用をサービス化すれば、依頼者が操作履歴データを用いて検証経過を辿る（再生する）ことによって、その正当性を確認すると共に

分析ノウハウを修得する新たなサービスを提供することもできる。

【0076】

なお、上記の受託者が行う帰結の取りまとめの際に行われる帰結の評価を、前述した図9のステップ1002の類似度判定を用いることで自動的に行うこともできる。

【0077】

帰結の評価には、

(a) 帰結が類似し、操作履歴が類似している（同じ分析内容である可能性が非常に高い）

(b) 帰結が類似しないが、操作履歴が類似している（同じ分析内容であるが、帰結の表現が異なる可能性が高い）

(c) 帰結が類似するが、操作履歴が類似していない（同じ帰結が、異なる分析過程により支持される可能性が高い）

(d) 帰結が類似せず、操作履歴も類似していない（双方が互いに異なる分析内容である）

の、4種類に分けられる。

【0078】

通常は、受託者が帰結の内容を参照して上記4種類の評価に分類するが、履歴比較部17で類似度判定を行って自動的に帰結の評価を分類することができる。

【0079】

分析者の帰結は、依頼者が理解可能なように、自然言語で記述されるのが一般的である。受託者は、これらの帰結を評価しようとする場合、同じ分析内容に同じ評価を与える必要がある。しかし、自然言語には、表現のゆれがあり、帰結の情報のみで分類すると精度が悪くなる。そこで、帰結の比較結果と、操作履歴の比較結果を前述したように併用し帰結を分類することで、評価作業を効率化することができる。

【0080】

以上説明したように第1の実施の形態の情報マイニングシステムでは、複数の分析者が分析を行っている途中の一連の操作を分析中操作履歴テーブルに記録し

、これを、既に登録された帰結及び該帰結に対応する分析の一連の操作を対応付けた操作履歴テーブルと逐次比較して類似度を判定し、類似度と所定の閾値（トリガー値）との比較結果に基づいて、操作者に警告するので、複数の分析者が同時並行的にマイニングを進める際に、重複した分析を防ぐことが可能となる。

【0081】

次に、本発明の第2の実施の形態について図面を参照して説明する。

【0082】

第2の実施の形態では、第1の実施の形態と比較して、分析者側の分析者コンピュータ4に分析ツール14-a、14-b、が備えられ、分析用データ、操作履歴、帰結等のデータ及び情報を、ネットワーク5-cを介して、情報マイニングシステム1と受け渡しを行う点が異なる。なお、第1の実施の形態と同一の動作をする構成には同一の符号を付し、その説明は省略する。

【0083】

図12は第2の実施の形態の情報マイニングシステムを中心とした構成を示すブロック図である。

【0084】

情報マイニングシステム1には、分析ツールは備えられておらず、分析者のコンピュータ4-a及び4-bには、それぞれ分析ツール14-a及び14-bが備えられている。

【0085】

なお、この分析ツール14は、最初から分析者コンピュータ4に備えられていてもよいし、情報マイニングシステム1の記憶部に格納し、分析ID通知時に分析者コンピュータ4にダウンロードするような仕組みとしてもよい。

【0086】

次に、第2の実施の形態の情報マイニングシステムの動作を、図2において前述したように、8つのステップに分けて説明する。

【0087】

図13は、第2の実施の形態の情報マイニングシステムの処理を示すシーケンス図である。

【0088】

なお、図2で説明したのと同様に受託者は情報マイニングシステム1の管理者であり、受託者と情報マイニングシステムは一つに示してある。

【0089】

まず、第1のステップでは、依頼者が情報マイニングシステムの受託者に対して、データマイニングを要求し、マイニングを委託する。

【0090】

次に、第2のステップでは、依頼者が受託者に対して分析対象データベース6に保持されたデータの情報を通知して、情報マイニングシステム1から分析対象データベース6に接続可能にし（シーケンス1）、分析対象データベース6のデータ内容をネットワーク5-aを経由して情報マイニングシステム1に転送する。また、データの転送が認められない場合には、依頼者のコントロールの下ネットワーク5-aを介した、情報マイニングシステム1から分析対象データベース6への接続を確立し、分析対象データへのアクセスを可能とする。なお、記録媒体に記録した分析対象データを依頼者から受託者に受け渡してもよい。

【0091】

次に、第3のステップは、情報マイニングシステム1の分析用データ生成部11によって、分析対象データベース6の分析対象データを分析ツール14で使用可能なデータ形式に変換し、分析用データを生成して、分析用データ保持部12に保存する。

【0092】

次に、第4のステップは、受託者が、複数の分析者（分析者コンピュータ4）に対して、分析ID（分析ツール14へのアクセスに関する情報、例えば、分析ツール14を使用するためのユーザID等）を通知し、分析用データのマイニングを依頼する（シーケンス2-a）。なお、分析者（分析者コンピュータ4）は複数でなくてもよい。

【0093】

次に、第5のステップでは、分析データのマイニング依頼を受けた分析者が、分析者コンピュータ4において分析ツール14を起動し、分析ツール14に対し

て分析 ID をログインコマンドとして入力することで、分析ツール 14 にログインする。分析 ID を入力することで分析ツール 14 が利用可能となり、データの分析を開始できる。この際、分析ツール 14 は、分析の操作に従って、ネットワーク 5-c を経由して情報マイニングシステム 1 に記憶されている分析用データを要求する（シーケンス 3-a）。

【0094】

情報マイニングシステム 1 は、分析用データ保持部 12 から要求を受けた分析用データを分析者のコンピュータ（分析ツール 14）に対して送信する（シーケンス 4-a）。なお、分析ツール 14 が情報マイニングシステム 1 との間でデータ、メモ、帰結等の情報を送る際には、ステップ 4 で通知された分析 ID と共に送る。この分析 ID によって、どの分析者からの操作であるかを情報マイニングシステム 1 が認識することができる。

【0095】

各分析者は、分析データをブラウズ（参照）し、何らかの仮説を立案して、分析ツール 14 を使用して仮説を検証する。この際に、分析ツール 14 のメモ欄にメモを電子的に書き留め、記録することができる。このメモは、分析ツールの処理の履歴と共にネットワーク 5-c を介して情報マイニングシステム 1 の操作履歴データ保持部 15 に逐次記録される。さらに、仮説の検証の結果を帰結やコメントをメモ欄に記入し登録することができる。例えば、仮説の検証に成功し何らかの有益な知識が得られた場合にはその知識を登録し、仮説の検証に失敗した場合にはその失敗を登録し、検証をあきらめた場合にはあきらめた事実を登録することができる（シーケンス 5-a、9-a）。

【0096】

分析者の分析作業中は、分析者が分析ツール 14 に対して行った操作の履歴が、分析者毎に、そして分析処理が行われる毎に、ネットワーク 5-c を経由して情報マイニングシステム 1 の操作履歴データ保持部 15 に逐次記録される。また、分析者が仮説の検証の結果を記入したメモを、帰結として登録することを指示した場合には、当該メモが「帰結」となる。分析ツール 14 は、帰結の内容を、帰結に至った一連の検証の操作履歴と対応付けて、ネットワーク 5-c を経由し

て情報マイニングシステム 1 の帰結データ保持部 1 6 に記録する。

【0 0 9 7】

なお、帰結の登録が行われるまで、分析中の操作履歴は、操作履歴データ保持部 1 5 の分析中操作履歴テーブルに一時的に格納されている。そして、帰結を登録すると、操作履歴は帰結と対応付けられ、操作履歴データ保持部 1 5 の操作履歴テーブルに記録される。操作履歴テーブルに記録されると、対応する操作履歴は分析中操作履歴テーブルから削除される。つまり操作履歴は移動される。

【0 0 9 8】

次に、第 6 のステップでは、情報マイニングシステム 1 は、現在分析中の分析者が行ったネットワーク 5 - c を介して送られる操作履歴が追加される都度に、履歴比較部 1 7 によって、分析者の現在の分析に関わる操作履歴と、操作履歴テーブルに記録されている過去の操作履歴とを比較する。履歴比較部 1 7 は、分析者の分析中の操作履歴が他の分析者（又は、分析者自身が過去に行った分析）によって登録された帰結に対応する操作履歴（又はその一部）と類似度が高いと判定した場合、ネットワーク 5 - c を介して分析ツール 1 4（分析者自身の分析者コンピュータ 4）に対して警告を通知すると共に、類似度が高い操作履歴に対応する帰結やメモの一覧等を表示する（シーケンス 7）。この警告を受けた分析者は、警告に対する確認操作を行う（シーケンス 8 - a）。

【0 0 9 9】

分析者は、一覧を確認し自分が立案した仮説に対する帰結が、既に他の分析者（又は自分が過去に行った分析結果）によって得られているときは、重複した探索を防ぐために作業を中断することができる。

【0 1 0 0】

また、帰結が既に得られている場合においても作業を続行することもできる。これは、提示された時点までの操作履歴が類似していても、以降の分析で異なる帰結を得られたり、新たな知識の発見につながる事象を分析者が見出す可能性もあるからである。この場合、仮に同じ帰結に至った場合には、情報マイニングの多様性の点からは価値が低いだが、帰結の確度を補強する情報となる。この履歴比較部 1 7 による類似度の判定は、分析対象の類似性、分析条件（検索キーワード

や数値)の類似性、操作および操作の流れの類似性を数値化することによって求めることができる。

【0101】

前述した第5及び6のステップは繰り返し行われる。そして、第7のステップでは、予め定められた一定の期間(例えば、依頼者の指定した納期)が経過した場合、又は、予め定められた帰結の数(例えば、依頼者の指定したコストに基づく成果)が得られた場合に、分析者は分析を終了し、分析ツール14からのログアウト処理を行う。この分析者のログアウト又は分析者からの終了の通知は分析ツール14からネットワーク5-cを介して情報マイニングシステム1に伝えられ、一連の分析処理が終了する(シーケンス10-a)。

【0102】

そして、受託者は分析者の分析結果が報告した帰結の一覧を帰結データ保持部16から得る。この帰結の一覧は受託者が取りまとめを行う。この取りまとめは、例えば、表記や表現は異なるが同じ内容を示すと考えられる帰結を一つにまとめたり、得られた帰結の内容を勘案して評価付けを行う。

【0103】

なお、この後の受託者と依頼者とで行われる第8のステップは、第1の実施の形態の図11において説明したものと同一であるため説明を省略する。

【0104】


以上説明したように第2の実施の形態の情報マイニングシステムでは、分析者コンピュータ4に分析ツール14を備えたので、第1の実施の形態の効果に加え、情報マイニングシステム1の処理の負荷を軽減することができる。

【0105】

次に、本発明の第3の実施の形態について図面を参照して説明する。

【0106】

第3の実施の形態では、第1及び第2の実施の形態と比較して、分析者側の分析者コンピュータ4に分析ツール14-a、14-bが備えられ、さらに、分析者コンピュータ4において分析用データの内容の全て又は一部を保持している点異なる。なお、第1又は第2の実施の形態と同一の動作をする構成には同一の



符号を付し、その説明は省略する。

【 0 1 0 7 】

図 1 4 は第 3 の実施の形態の情報マイニングシステムを中心とした構成を示すブロック図である。

【 0 1 0 8 】

情報マイニングシステム 1 には、分析ツールは備えられておらず、分析者のコンピュータ 4 - a 及び 4 - b には、それぞれ分析ツール 1 4 - a 及び 1 4 - b が備えられている。

【 0 1 0 9 】

なお、この分析ツール 1 4 は、最初から分析者コンピュータ 4 に備えられていてもよいし、情報マイニングシステム 1 の記憶部に格納し、分析 I D 通知時に分析者コンピュータ 4 にダウンロードするような仕組みとしてもよい。

【 0 1 1 0 】

さらに、分析者コンピュータ 4 - a 及び 4 - b には、それぞれ分析用データのコピーである分析用データ保持部 1 2 - a 及び 1 2 - b が備えられる。このデータは、分析者が分析を開始する際に分析 I D と共に情報マイニングシステム 1 から送られる分析用データの概要を参照して分析者（又は、分析者コンピュータ 4）が、分析用データのどの部分（範囲）を受け取り保持するかを決定する。決定した分析用データの範囲の情報を情報マイニングシステム 1 に対して通知する。情報マイニングシステム 1 は、分析者コンピュータ 4 から送られた分析用データの範囲の情報を受け取り、分析用データ保持部 1 2 から該当する範囲の分析用データを分析者コンピュータ 4 に対して送る。なお、情報マイニングシステム 1 が、予め分析者コンピュータ 4 毎に分析用データの範囲を決定し、それぞれの範囲の分析用データを分析者コンピュータ 4 に対して送ってもよい。また、各々の分析者コンピュータ 4 が分析用データのすべての範囲を持つようにしてもよい。

【 0 1 1 1 】

次に、第 3 の実施の形態の情報マイニングシステムの動作を、図 2 において前述したように、8 つのステップに分けて説明する。

【 0 1 1 2 】

図15は、第3の実施の形態の情報マイニングシステムの処理を示すシーケンス図である。

【0113】

なお、図2で説明したのと同様に受託者は情報マイニングシステム1の管理者であり、受託者と情報マイニングシステムは一つに示してある。

【0114】

まず、第1のステップでは、依頼者が情報マイニングシステムの受託者に対して、データマイニングを要求して、マイニングを委託する。

【0115】

次に、第2のステップでは、依頼者が受託者に対して分析対象データベース6に保持されたデータの情報を通知して、情報マイニングシステム1から分析対象データベース6に接続可能にし（シーケンス1）、分析対象データベース6のデータ内容をネットワーク5-aを経由して情報マイニングシステム1に転送する。また、データの転送が認められない場合には、依頼者のコントロールの下ネットワーク5-aを介した、情報マイニングシステム1から分析対象データベース6への接続を確立し、分析対象データへのアクセスを可能とする。なお、記録媒体に記録した分析対象データを依頼者から受託者に受け渡してもよい。

【0116】

次に、第3のステップは、情報マイニングシステム1の分析用データ生成部11によって、分析対象データベース6の分析対象データを分析ツール14で使用可能なデータ形式に変換し、分析用データを生成して、分析用データ保持部12に保存する。

【0117】

次に、第4のステップは、受託者が、複数の分析者（分析者コンピュータ4）に対して、分析ID（分析ツール14へのアクセスに関する情報、例えば、分析ツール14を使用するためのユーザID等）、及び、分析用データの概要を通知し、分析用データのマイニングを依頼する（シーケンス2-b）。この分析用データの概要は、分析用データ保持部12に保持されている分析用データの概要である。

【0118】

分析者（又は、分析者コンピュータ 4）は、受け取った分析用データの概要から分析用データのどの部分（どの範囲）を受け取り保持するかを決定し、決定した分析用データの範囲の情報を、情報マイニングシステム 1 に対して要求する（シーケンス 3-b）。なお、分析者（分析者コンピュータ 4）は複数でなくともよい。

【0119】

情報マイニングシステム 1 は、分析者コンピュータ 4 から送られた分析用データの範囲の情報を受け取り、分析用データ保持部 12 から該当する範囲の分析用データを分析者コンピュータ 4 に対して送る（シーケンス 4-b）。

【0120】

次に、第 5 のステップでは、分析データのマイニング依頼を受けた分析者が、分析者コンピュータ 4 において分析ツール 14 を起動し、分析ツール 14 に対して分析 ID をログインコマンドとして入力することで、分析ツール 14 にログインする。分析 ID を入力することで分析ツール 14 が利用可能となり、分析用データ保持部 12-a 又は 12-b に保持されたデータの分析を開始できる。

【0121】

分析ツール 14 は、分析者コンピュータ 4 に保持されている分析用データを参照する。各分析者は、分析データをブラウズ（参照）し、何らかの仮説を立案して、分析ツール 14 を使用して仮説を検証する。この際に、分析ツール 14 のメモ欄にメモを電子的に書き留め、記録することができる。このメモは、分析ツールの処理の履歴と共にネットワーク 5-c を介して情報マイニングシステム 1 の操作履歴データ保持部 15 に逐次記録される。さらに、仮説の検証の結果やコメントをメモ欄に記入し登録することができる。例えば、仮説の検証に成功し何らかの有益な知識が得られた場合にはその知識を登録し、仮説の検証に失敗した場合にはその失敗を登録し、検証をあきらめた場合にはあきらめた事実を登録することができる（シーケンス 5-b、9-b）。

【0122】

なお、分析ツール 14 が情報マイニングシステム 1 との間でデータ、メモ、仮

説等の情報を送受信する際には、ステップ4で通知された分析IDと共に送る。この分析IDによって、どの分析者からの操作であるかを情報マイニングシステム1が認識することができる。

【0123】

分析者の分析作業中は、分析者が分析ツール14に対して行った操作の履歴が、分析者毎に、そして分析処理が行われる毎に、ネットワーク5-cを経由して情報マイニングシステム1の操作履歴データ保持部15に逐次記録される。また、分析者が仮説の検証の結果を記入したメモを、帰結として登録することを指示した場合には、当該メモが「帰結」となる。分析ツール14は、帰結の内容を、帰結に至った一連の検証の操作履歴と対応付けて、ネットワーク5-cを経由して情報マイニングシステム1の帰結データ保持部16に記録する。

【0124】

なお、帰結の登録が行われるまで、分析中の操作履歴は、操作履歴データ保持部15の分析中操作履歴テーブルに一時的に格納されている。そして、帰結を登録すると、操作履歴は帰結と対応付けられ、操作履歴データ保持部15の操作履歴テーブルに記録される。操作履歴テーブルに記録されると、対応する操作履歴は分析中操作履歴テーブルから削除される。つまり操作履歴は移動される。

【0125】

次に、第6のステップでは、情報マイニングシステム1は、現在分析中の分析者が行ったネットワーク5-cを介して送られる操作履歴が追加される都度に、履歴比較部17によって、分析者の現在の分析に関わる操作履歴と、操作履歴テーブルに記録されている過去の操作履歴とを比較する。履歴比較部17は、分析者の分析中の操作履歴が他の分析者（又は、分析者自身が過去に行った分析）によって登録された帰結に対応する操作履歴（又はその一部）と類似度が高いと判定した場合、ネットワーク5-cを介して分析ツール14（分析者自身の分析者コンピュータ4）に対して警告を通知すると共に、類似度の高い操作履歴に対応する帰結やメモの一覧等を表示する。（シーケンス7）。この警告を受けた分析者は、警告に対する確認操作を行う（シーケンス8-b）。

【0126】

分析者は、一覧を確認し自分が立案した仮説に対する帰結が、既に他の分析者（又は自分が過去に行った分析結果）によって得られているときは、重複した探索を防ぐために作業を中断することができる。

【0 1 2 7】

また、帰結が既に得られている場合においても作業を続行することもできる。これは、提示された時点までの操作履歴が類似していても、以降の分析で異なる帰結を得られたり、新たな知識の発見につながる事象を分析者が見出す可能性もあるからである。この場合、仮に同じ帰結に至った場合には、情報マイニングの多様性の点からは価値が低いが、帰結の確度を補強する情報となる。この履歴比較部 1 7 による類似度の判定は、分析対象の類似性、分析条件（検索キーワードや数値）の類似性、操作および操作の流れの類似性を数値化することによって求めることができる。

【0 1 2 8】

前述した第 5 及び第 6 のステップは繰り返し行われる。そして、第 7 のステップでは、予め定められた一定の期間（例えば、依頼者の指定した納期）が経過した場合、又は、予め定められた帰結の数（例えば、依頼者の指定したコストに基づく成果）が得られた場合に、分析者は分析を終了し、分析ツール 1 4 からのログアウト処理を行う。この分析者のログアウト又は分析者からの終了の通知は分析ツール 1 4 からネットワーク 5 - c を介して情報マイニングシステム 1 に伝えられ、一連の分析処理が終了する（シーケンス 1 0 - b）。なお、セキュリティの観点から、分析ツール 1 4 の終了時に分析者コンピュータ 4 に保持された分析用データを破棄することが望ましい。

【0 1 2 9】

そして、受託者は分析者の分析結果が報告した帰結の一覧を帰結データ保持部 1 6 から得る。この帰結の一覧は受託者が取りまとめを行う。この取りまとめは、例えば、表記や表現は異なるが同じ内容を示すと考えられる帰結を一つにまとめたり、得られた帰結の内容を勘案して評価付けを行う。

【0 1 3 0】

なお、この後の受託者と依頼者とで行われる第 8 のステップは第 1 の実施の形

態の図 1 1 において説明したものと同一であるため説明を省略する。

【0 1 3 1】

以上説明したように第 3 の実施の形態の情報マイニングシステムでは、分析者コンピュータ 4 に分析用データの内容の少なくとも一部保持したので、第 1 及び第 2 の実施の形態の効果に加え、ネットワーク（ネットワーク 5 - c）で送受信されるデータの量を軽減し、ネットワークのトラフィックを低減することができる、情報マイニングシステム全体としての処理の効率を上げることができる。

【0 1 3 2】

【発明の効果】

本発明の情報マイニングシステムでは、複数の分析者が同時並行的にマイニングを進める際に、複数分析者の使用する端末の各々が行った分析ツールの操作履歴と、分析によって得られた帰結と、を対応付けて記録する帰結データ保持部と、分析中の複数の端末の行っている分析ツールの一連の操作の履歴を記録する操作履歴データ保持部と、帰結データ保持部の内容の少なくとも一部が、操作履歴データ保持部の内容に一致又は類似するか否かを判定する履歴比較部とを備え、履歴比較部によって帰結データ保持部の内容と、操作履歴データ保持部との内容が一致又は類似すると判定された場合に、前記分析中の分析ツールに対して通知を行うので、複数の分析者が重複した探索を防ぐことができる。また、警告が発せられた際に、他の分析者によって既に得られた帰結一覧とその履歴情報を参照する手段を設けることで、効率よく多様な知識発見を行うことができる。さらに、履歴比較部を用いて帰結と操作履歴を比較併用することで、帰結の評価作業を効率化できるという効果がある。

【図面の簡単な説明】

【図 1】 本発明の第 1 の実施の形態の、情報マイニングシステムを中心とした構成を示すブロック図である。

【図 2】 同じく、情報マイニングシステムの処理のシーケンス図である。

【図 3】 同じく、関連語を表示する関連語表示ツールの画面表示の説明図である。

【図 4】 同じく、指定した文字列を含む文章を検索する全文検索ツールの

画面表示の説明図である。

【図 5】 同じく、指定した文章と類似する文章を検索する類似文検索ツールの画面表示の説明図である。

【図 6】 同じく、分析中操作履歴テーブルの説明図である。

【図 7】 同じく、操作履歴テーブルの説明図である。

【図 8】 同じく、帰結テーブルの説明図である。

【図 9】 同じく、操作履歴の類似度判定処理のフローチャートである。

【図 10】 同じく、警告の通知時に分析者コンピュータ 4 に表示される画面の説明図である。

【図 11】 同じく、分析された結果を受託者が依頼者に対して提示する処理のシーケンス図である。

【図 12】 本発明の第 2 の実施の形態の、情報マイニングシステムを中心とした構成を示すブロック図である。

【図 13】 同じく、情報マイニングシステムの処理のシーケンス図である。

【図 14】 本発明の第 3 の実施の形態の、情報マイニングシステムを中心とした構成を示すブロック図である。

【図 15】 同じく、情報マイニングシステムの処理のシーケンス図である。

【符号の説明】

- 1 情報マイニングシステム
- 2 依頼者コンピュータシステム
- 3 受託者コンピュータ
- 4、4-a、4-b 分析者コンピュータ
- 5-a、5-b、5-c ネットワーク
- 6 分析対象データベース
- 11 分析用データ生成部
- 12、12-a、12-b 分析用データ保持部
- 13 一般情報データ保持部

1 4、1 4 - a、1 4 - b 分析ツール

1 5 操作履歴データ保持部

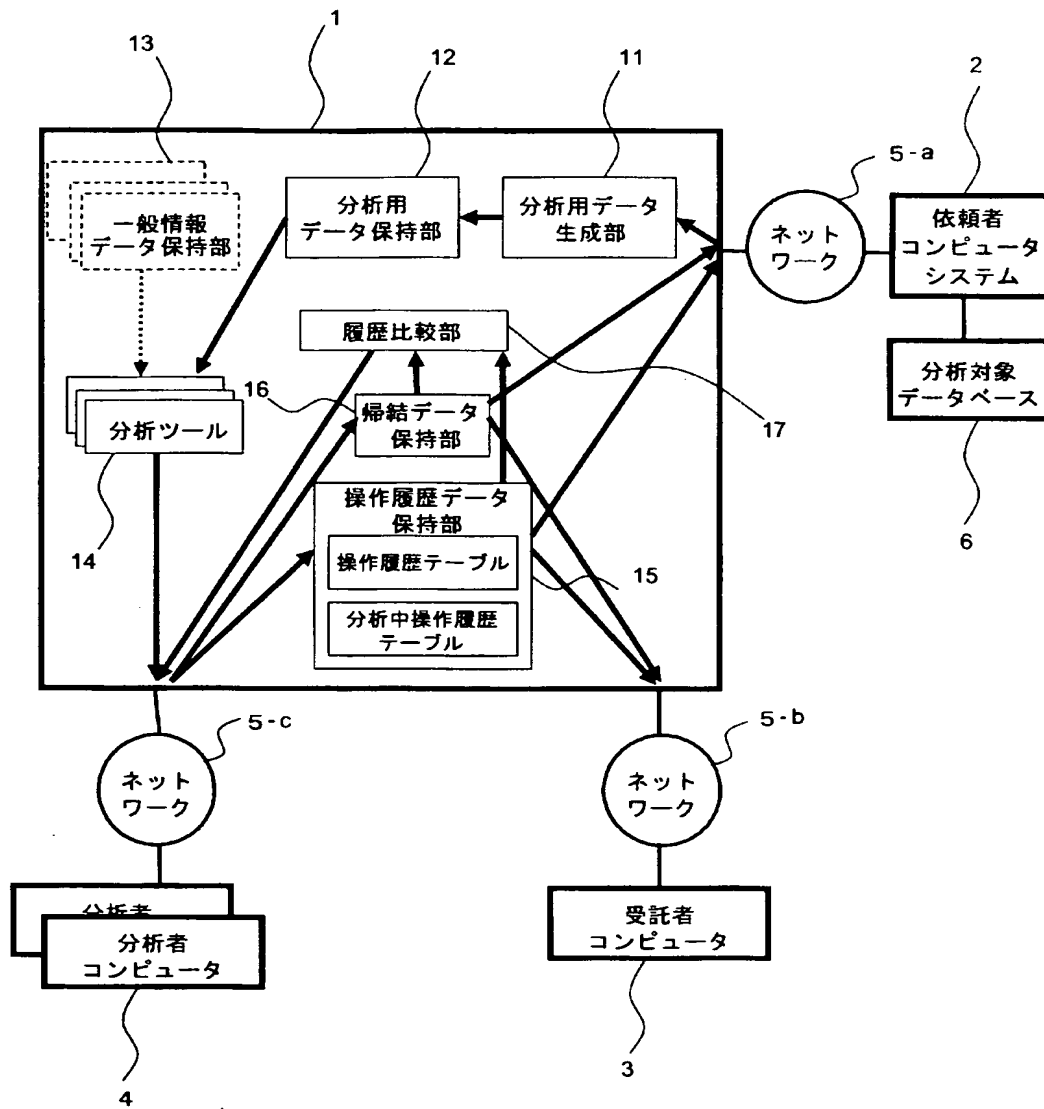
1 6 帰結データ保持部

1 7 履歴比較部

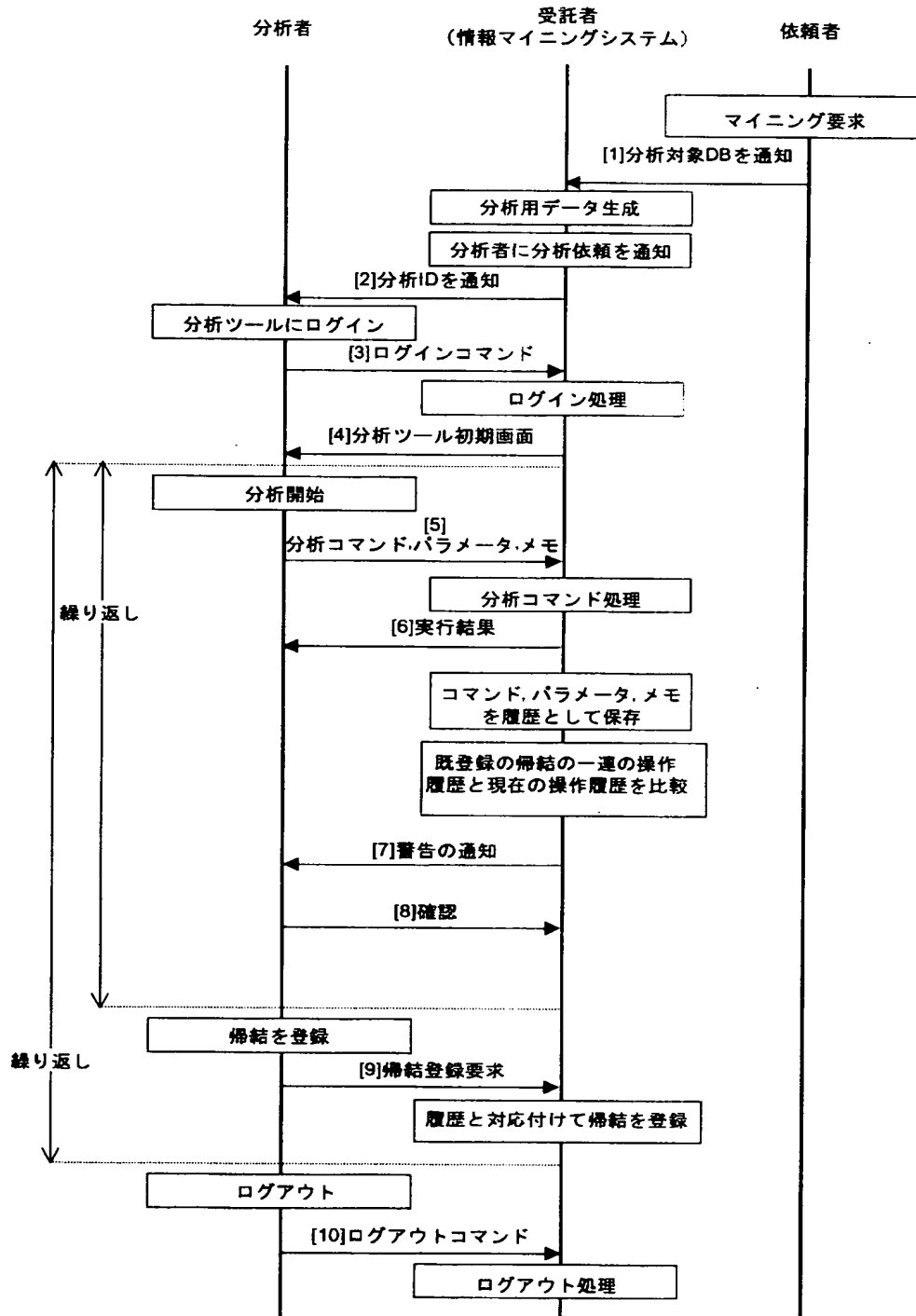
【書類名】

図面

【図 1】



【図2】



【図 3】

関連語表示ツール

単語頻度一覧表示

実行

単語共起語一覧表示

実行

入力語

ヒューズ, 配線

表示形式

☐ グループ

☒ ペア

データ条件設定

メモ

帰結登録

結果表示

単語共起語一覧表示

入力語	関連語	頻度
ヒューズ	切れる	100
ヒューズ	交換	70
配線	不良	50
ヒューズ	切断	10
配線	切断	5

【図 4】

全文検索ツール

全文検索

検索語

ヒューズ, 切れる

条件

☒すべてを含む(AND)
☐いずれかを含む(OR)

データ条件設定

メモ

実行

帰結登録

結果表示

検索結果一覧

文書番号	内容
15	アイロンのヒューズが切れたがどのように交換すればいいのか
58	ヒューズがよく切れる。配線に問題があるのではないか

【図 5】

類似文検索ツール

類似文検索
種文章

実行

アイロンのヒューズが切れたがどのように交換すればいいのか

データ条件設定

メモ

帰結登録

結果表示

検索結果一覧

文書 番号	類似 度	内容
15	100	アイロンのヒューズが切れたがどのように交換すればいいのか
112	65	アイロンのヒューズが切れたかどうか確認したい。
58	60	ヒューズがよく切れる。配線に問題があるのではないか
13	55	アイロンのコードを交換したい

【図 6】

記録 番号	対象データベース	分析操作種別	パラメータ
1	問合せデータ	単語頻度一覧	{期間=2001/6/1-2001/6/31}
2	問合せデータ	関連語一覧	{単語="ヒューズ"} & {期間=2001/6/1-2001/6/31}
3	問合せデータ	全文検索	{検索語=ヒューズ, 切れる} & {期間=2001/6/1- 2001/6/31}[コメント:増加傾向を調べる]

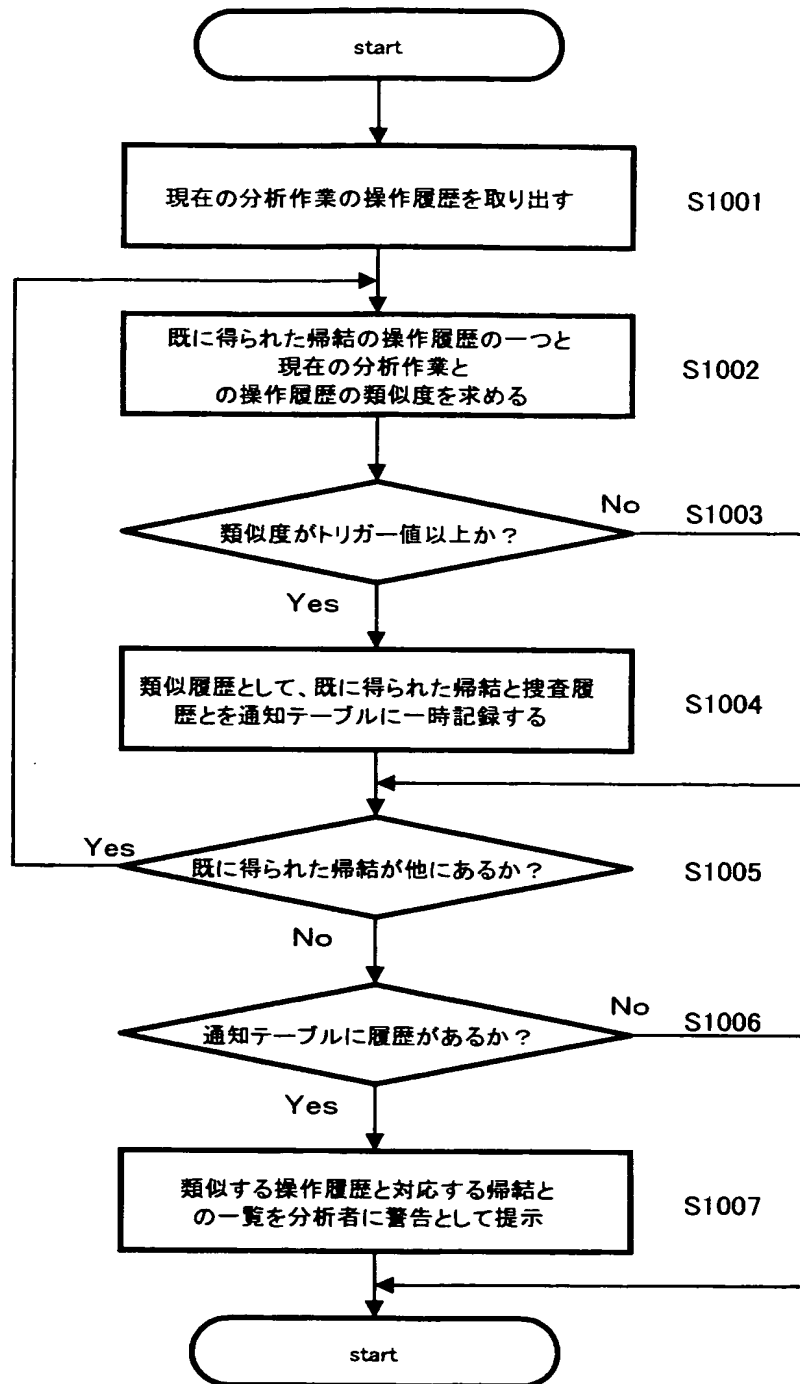
【図 7】

分析者ID	記録 番号	対象データベース	分析操作種別	パラメータ
:	:	:	:	:
TANAKA	21	問合せデータ	単語頻度一覧	{期間=2001/6/1-2001/6/31}
TANAKA	22	問合せデータ	関連語一覧	{単語="ヒューズ"} & {期間=2001/6/1-2001/6/31}
TANAKA	23	問合せデータ	全文検索	{検索語=ヒューズ, 切れる} & {期間=2001/6/1- 2001/6/31}[コメント: 増加.]
:	:	:	:	:
SUZUKI	115	問合せデータ	単語頻度一覧	{期間=2001/6/1-2001/6/31}

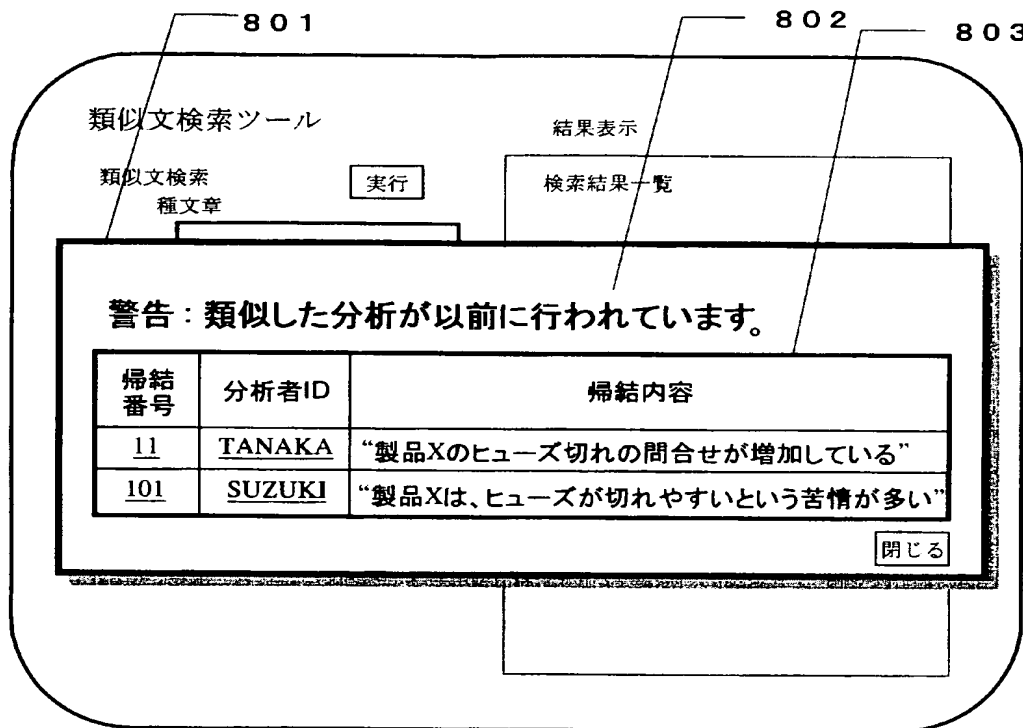
【図 8】

帰結 番号	分析者ID	対応履歴番号	帰結内容
11	TANAKA	21-23	“製品Xのヒューズ切れの問合せが増加 している”
101	SUZUKI	115-121	“製品Xは、ヒューズが切れやすいという 苦情が多い”

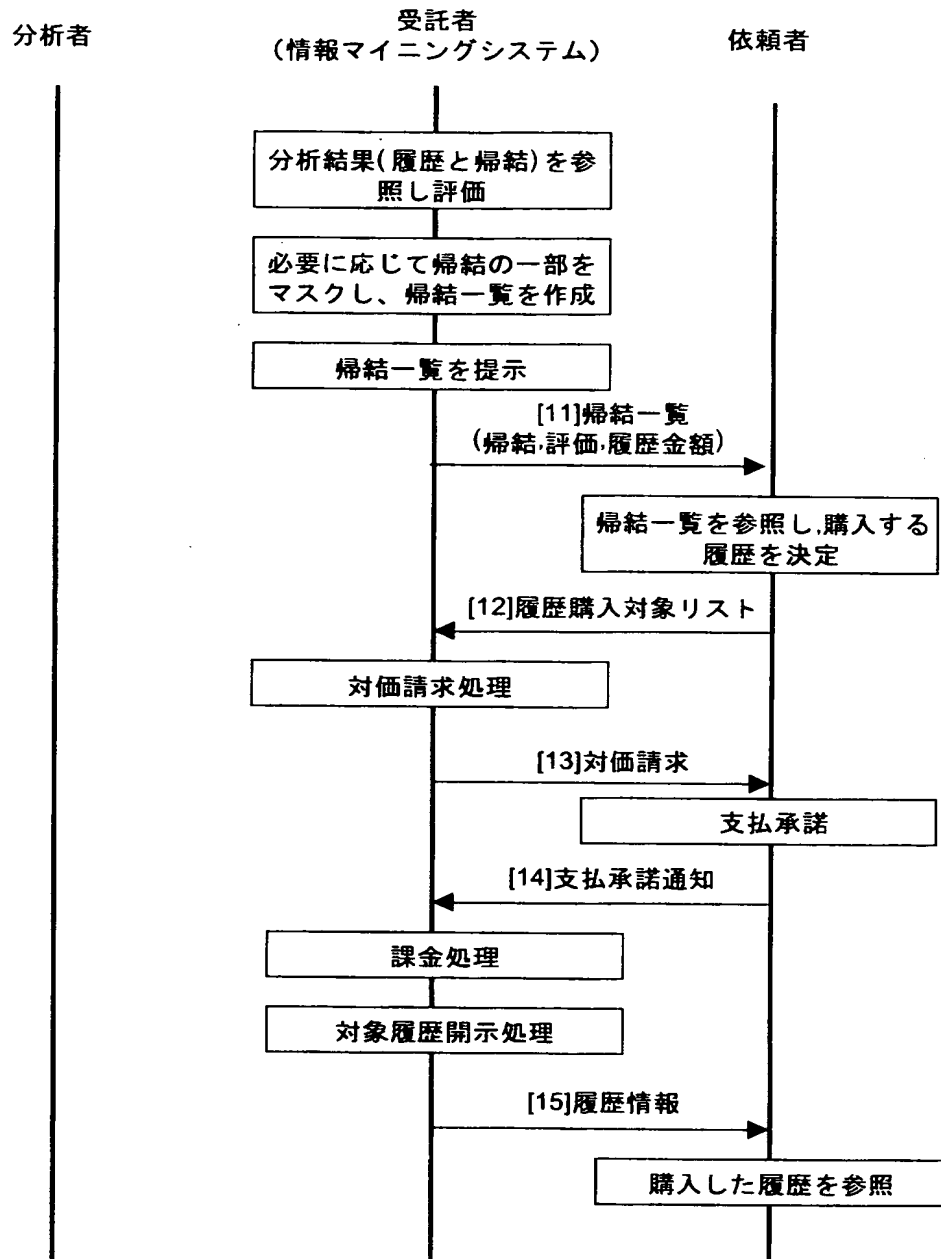
【図 9】



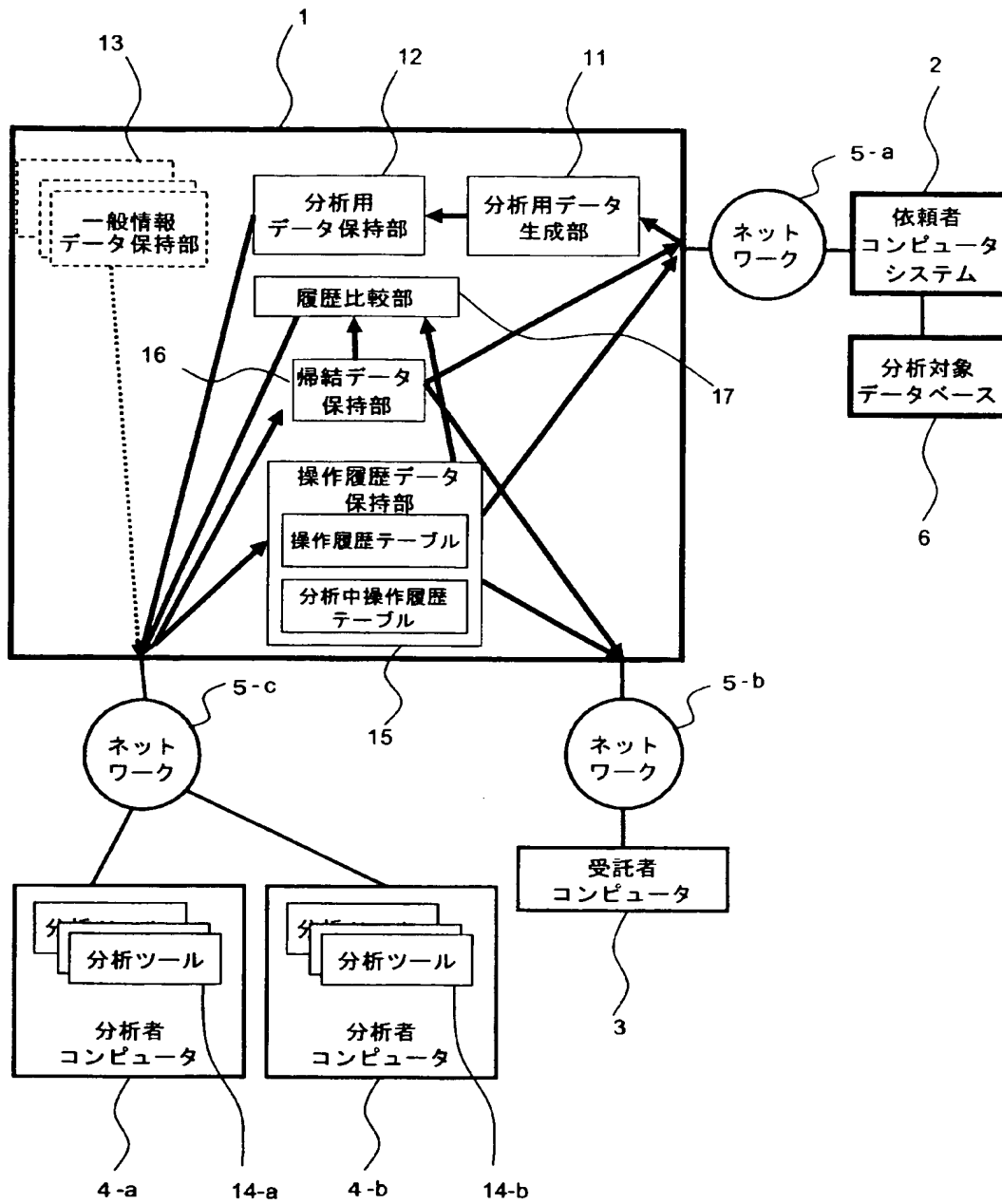
【図 10】



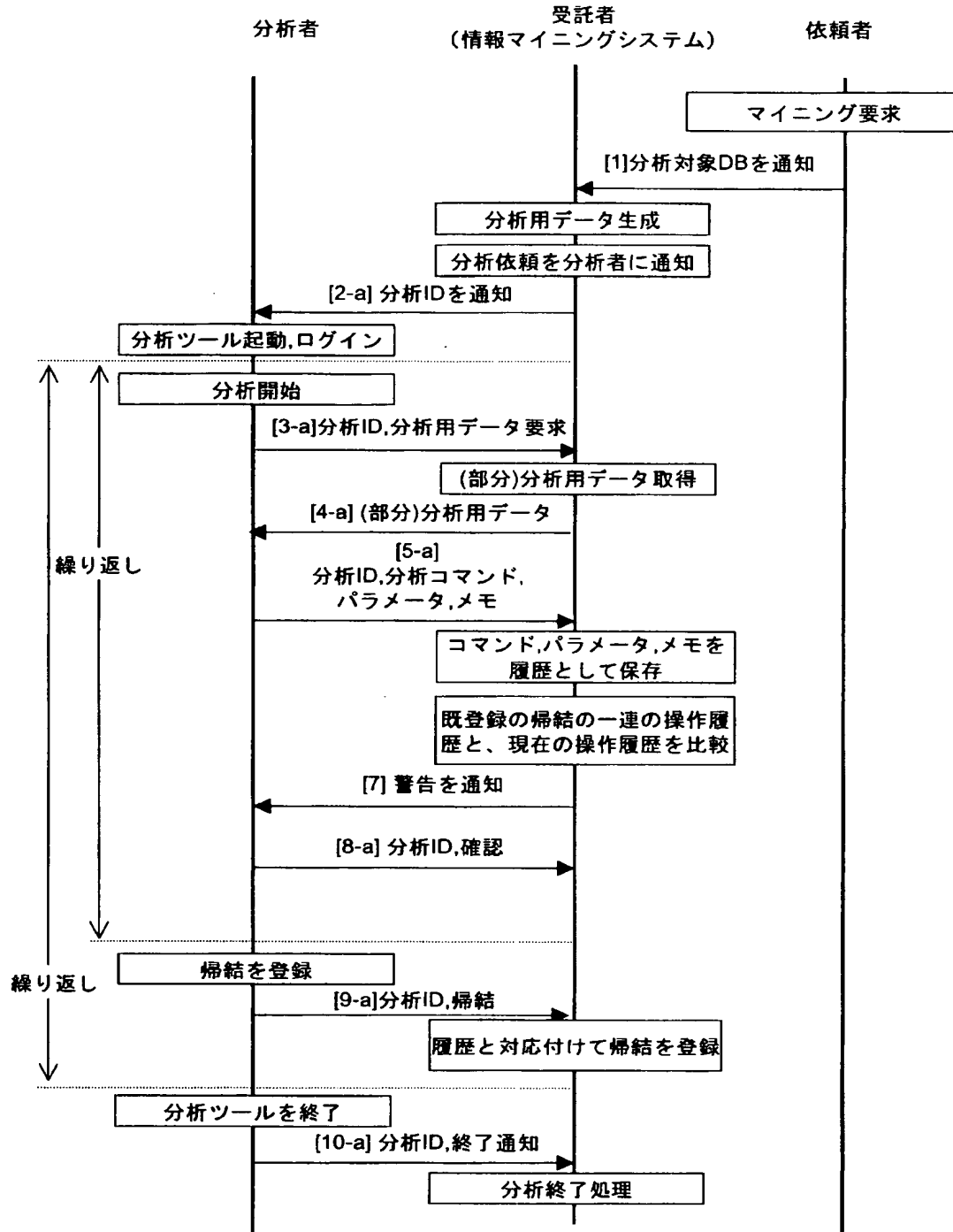
【図 1 1】



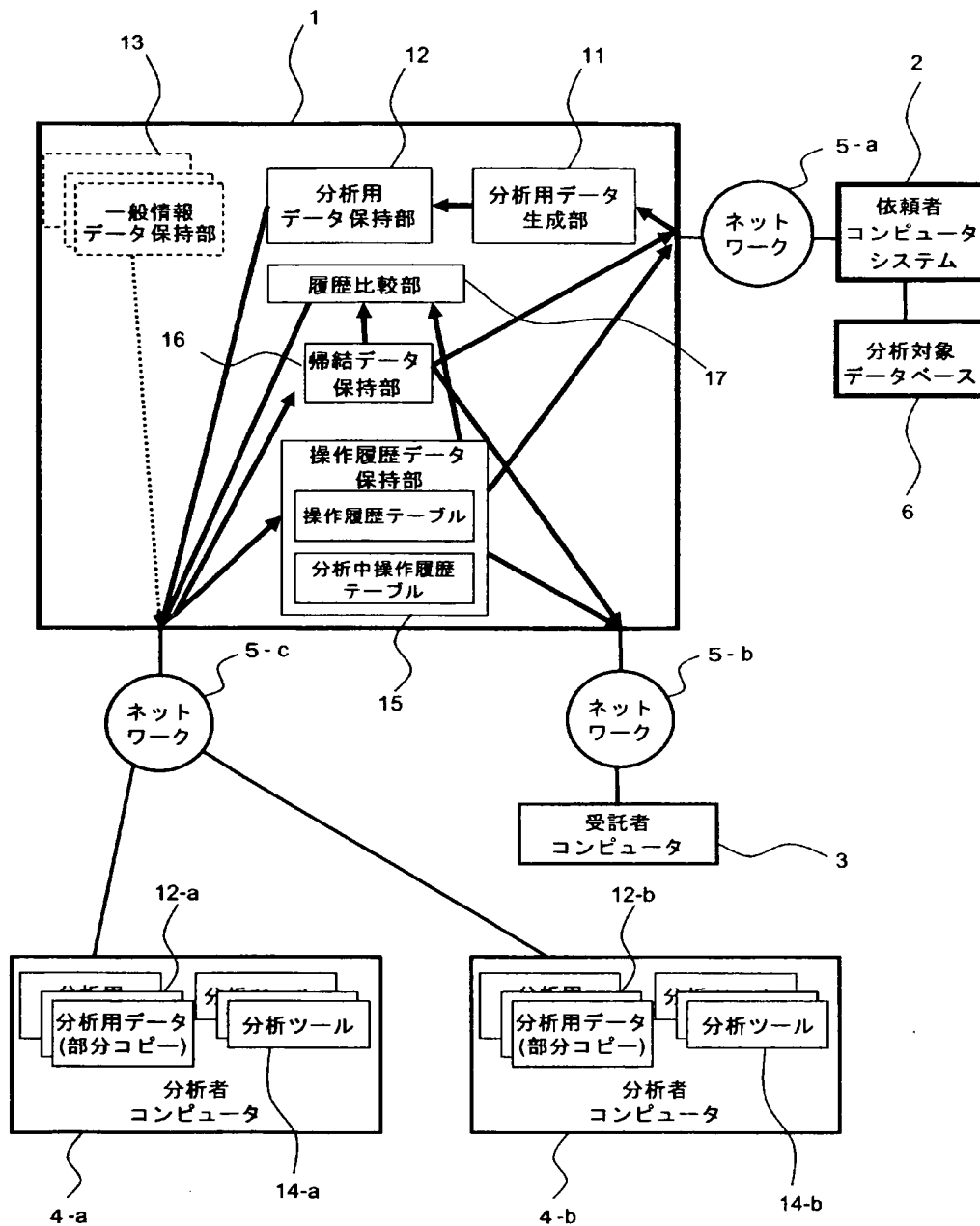
【図 12】



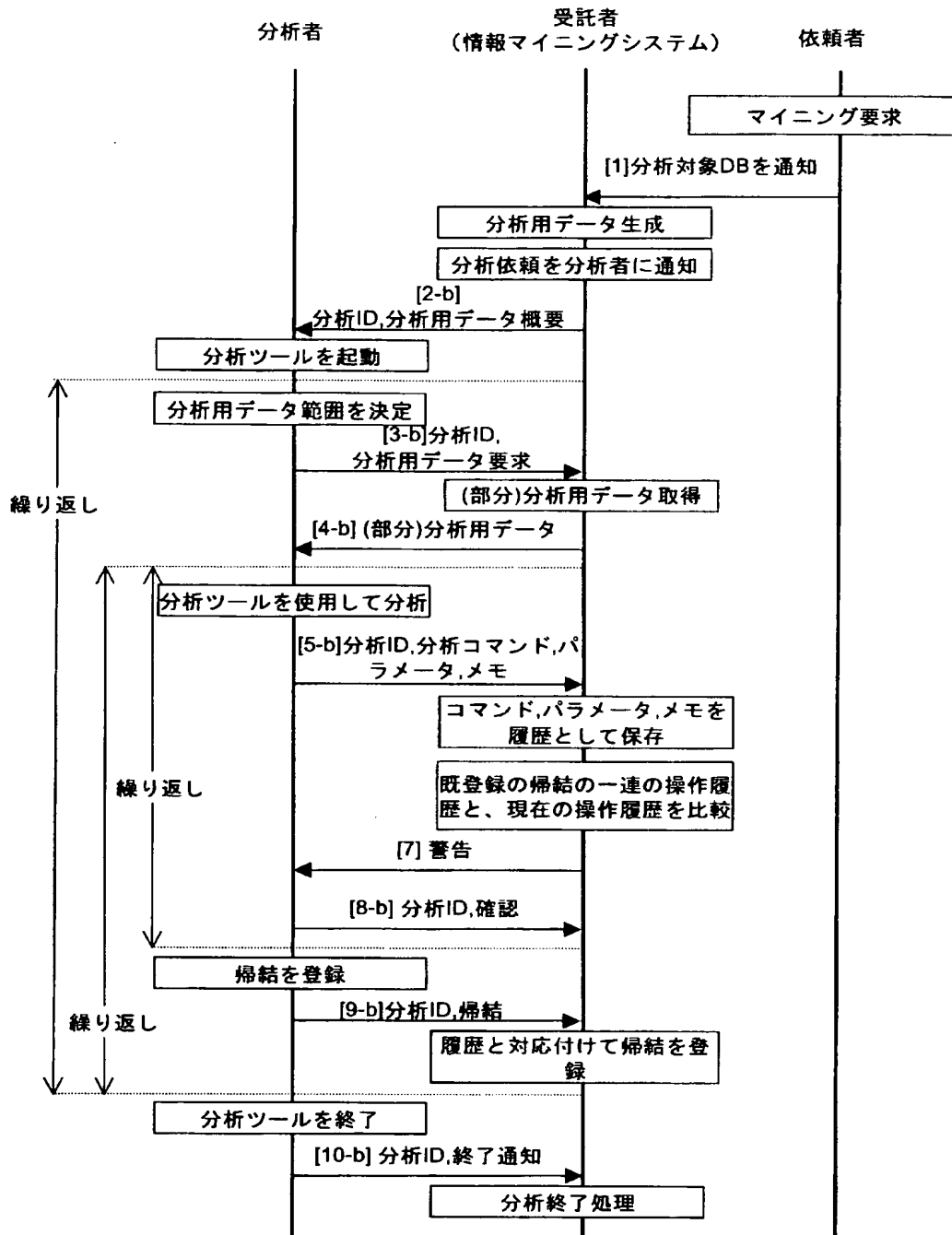
【図 13】



【図 14】



【図 15】



【書類名】 要約書

【要約】

【課題】

本発明は情報マイニングシステムに関し、特に、分析者が効率よくマイニングを行えるような情報マイニングシステムに関する。

【解決手段】

分析において、過去に行われた分析ツールの操作履歴及び分析中の分析ツールの操作履歴を記録する操作履歴データ保持部 15 と、過去に行われた分析ツールの操作履歴と分析中の分析ツールの操作履歴との類否を判定し、該過去に行われた分析ツールの操作履歴と該分析中の分析ツールの操作履歴の少なくとも一部とが類似すると判定された場合に、分析中の分析ツールに対してその旨を通知する履歴比較部 17 と、を備えた。

【選択図】 図 1

特願 2 0 0 3 - 1 5 9 2 7 4

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所